Lipid Fingerprints and Cofactor Dynamics of Light-Harvesting Complex II in Different Membranes

Thallmair, Sebastian; Vainikka, Petteri A.; Marrink, Siewert J.

Published in:
Biophysical Journal

DOI:
10.1016/j.bpj.2019.03.009

Publication date:
2019

Lipid fingerprints and cofactor dynamics of light-harvesting complex II in different membranes

Sebastian Thallmair[1]*, *Petteri A. Vainikka*[1,2], *Siewert J. Marrink*[1]*

[1]Groningen Biomolecular Sciences and Biotechnology Institute and The Zernike Institute for Advanced Material, University of Groningen, Nijenborgh 7, 9747 AG Groningen, Netherlands

[2]Department of Chemistry, University of Turku, 20014 Turku, Finland

RUNNING TITLE: LHCII dynamics and lipid fingerprints

KEYWORDS: photosynthesis – coarse-grained – molecular dynamics – thylakoid membrane – lipid fingerprints.

STATEMENT OF SIGNIFICANCE. Protein machineries involved in natural photosynthesis rely on multiple antenna complexes which capture the light energy and funnel it to the reaction center. Here, we provide coarse-grained simulations of light-harvesting complex II, the major antenna complex of plants. Our simulations reveal that the chromophores, which are responsible for the light capturing, change their flexibility and average distances upon trimer formation. This suggests that interactions with other proteins can impact the optical properties of the antenna complex. In addition, the lipid fingerprint of light-harvesting complex II shows a strong preference for monogalactosyldiacylglycerol lipids which are non-bilayer forming lipids present in the thylakoid membrane.

ABSTRACT. Plant light-harvesting complex II (LHCII) is the key antennae complex for plant photosynthesis. We present coarse-grained molecular dynamics simulations of monomeric and trimeric LHCII in a realistic thylakoid membrane environment based on the Martini force field. The coarse-grained protein model has been optimized with respect to atomistic reference simulations. Our simulations provide detailed insights in the thylakoid lipid fingerprint of LHCII which compares well with experimental data from membrane protein purification. Comparing the monomer and trimeric LHCII reveals a stabilizing effect of trimerization on the chromophores as well as the protein. Moreover, the average chromophore distance shortens in the trimer leading to stronger excitonic couplings. When changing the native thylakoid environment to a model membrane the protein flexibility remains constant, whereas the chromophore flexibility is reduced. Overall, the presented LHCII model lays the foundation to investigate the μs dynamics of this key antennae protein of plants.

1. INTRODUCTION

Plants and several bacteria exploit sunlight – the most abundant regenerative energy source – to generate energy. This highly developed and specified process – photosynthesis – takes place in the thylakoid membrane. Large protein supercomplexes collect the photons emitted by the sun and convert their energy in multiple steps into chemical energy which makes it more stable, storable and available at any time.

The photosystem II (PSII) supercomplex is involved in the first step in photosynthesis. PSII contains multiple antennae proteins which are responsible for the photon collection and the funneling of the excitation energy towards the reaction center. The latter performs the first conversion step by splitting water into molecular oxygen and protons. The generated electrons are stored in plastoquinole which is then further processed by cytochrome bf6.

In plants, the antenna system of PSII mainly consists of light-harvesting complex II (LHCII) associated to PSII in its trimeric form (1–3). In the case of high light exposure, the regulation of LHCII plays a key role in the photo protection mechanism. The acidified pH in the thylakoid lumen (1, 4) and a change of the composition of the carotenoid pool (1, 4, 5) are only two of the discussed factors. Altogether, LHCII has a crucial function in the first steps of photosynthesis and a more detailed microscopic picture of its dynamics will enhance our current understanding of how photosynthesis is regulated.

Here, we present a coarse-grained (CG) model at almost atomistic resolution of LHCII based on the Martini force field (6, 7). The Martini model is widely applied to study protein-lipid interactions in general (8–10), and recently also to reveal cofactor dynamics of photosystem II (11, 12). Our purpose is to provide microscopic insights on the behavior of LHCII in its natural environment on the hundreds of μs time scale. With our CG model, we aim to bridge the

description of physiologically important protein-protein interactions and supercomplex formation with the orientation dynamics of the key players in natural light harvesting, the chromophores. Both aspects are of prior necessity to be able to model the photophysical properties of LHCII and its regulation, defining the outstanding role of LHCII in natural light harvesting.

We constructed a CG model for a LHCII monomer which is benchmarked against recent atomistic molecular dynamics (MD) data (13). We analyze the trimer stability, changes upon monomer formation, the cofactor flexibility, as well as the impact of the membrane environment. Finally, we investigate the preferred lipid surrounding of trimeric LHCII, its lipid fingerprint. The presented CG model lays the foundation to a detailed investigation of the μs dynamics of the LHCII chromophores involved in light capturing.

The paper is structured as follows: The simulation details and the CG models together with their validations are given in Section 2. Section 3 focusses on the properties of LHCII trimers in thylakoid membrane and model membranes (3.1), the cofactor dynamics in both environments and their changes upon monomer formation (3.2), as well as the lipid fingerprints of LHCII (3.3). Finally, we conclude and give a brief outlook in Section 4.

2. METHODS

2.1 Simulation details

All MD simulations presented in our work have been performed using the program package Gromacs 2016.1 (14). All CG simulations were performed with the force field Martini 2.2 (6, 7). The simulation parameters were chosen in accordance to reference (15). In brief, the Verlet cutoff scheme was used with a buffer tolerance of 0.005 kJ/mol. Coulomb interactions were treated using the reaction-field method with a cutoff of 1.1 nm and a dielectric constant of $\varepsilon_r = 15$

for water. Van-der-Waals interactions were treated using the cut-off scheme with a cutoff of 1.1 nm. The velocity rescaling thermostat has been employed with a reference temperature of T = 310 K and a coupling constant of τ_T = 1 ps (16). For the equilibrations, the Berendsen barostat was employed (p = 1bar, τ_p = 3 ps) while the production runs were performed with a Parrinello-Rahman barostat (p = 1bar, τ_p = 12 ps) (17).

The CG simulations of LHCII have been performed in two different membrane environments: the native thylakoid membrane as well as a palmitoyloleoylglycerophosphocholine (POPC) model membrane. To enable a close comparison to the atomistic reference data (13) monomeric LHCII was embedded in a POPC bilayer consisting of 170 lipids per leaflet in a rectangular box (11.0×11.0×10.0 nm^3) using the program *insane.py* (18). The bilayer was solvated in ~5,600 CG water beads representing ~22,400 water molecules, neutralized, and 0.15 M NaCl was added. The optimization of the CG protein model was done including residues 13-231 resolved in the crystal structure, because the atomistic simulations only included these residues. CG simulations of 900 ns were used for the comparison with the reference data (Section 2.2). A detailed list of all simulations and the performed analysis is also given in Section S2 of the supporting information (SI).

To investigate the effect of trimer formation, trimeric as well as monomeric LHCII were embedded in a plant thylakoid membrane model (19), their natural membrane environment. Here, we use a simplified model for the plant thylakoid membrane based on the Martini model which has been characterized in reference (19). The plant thylakoid membrane model consists of four major lipid species: phosphatidylglycerol (PG), monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG), and sulfoquinovosyldiacylglycerol (SQDG) (20, 21). In combination with different lipid tails, the plant thylakoid membrane contains seven CG lipid

types – 16:1(3t)-16:0 PG, 16:1(3t)-18:3(9,12,15) PG, 18:3(9,12,15)-16:0 DGDG, di18:3(9,12,15) DGDG, 18:3(9,12,15)-16:0 MGDG, di18:3(9,12,15) MGDG, and 18:3(9,12,15)-16:0 SQDG – in the ratio of 14/28/14/100/14/72/42 (19). In case of trimeric LHCII, the thylakoid membrane contained in total 1188 lipids. It was solvated with ~21,900 CG water beads resulting in a box of 20.0×20.0×12.0 nm^3. Monomeric LHCII was embedded in a thylakoid membrane consisting of 328 lipids. It was solvated with ~4900 CG water beads resulting in a box of 11.0×11.0×10.0 nm^3. In both systems, 0.1 M NaCl was added after neutralization. CG simulations of 3 μs were used for the comparison of trimeric and monomeric LHCII (Sections 3.1 and 3.2). To obtain detailed information of the thylakoid lipid fingerprint (Section 3.3), two simulations of 30 μs each were performed.

To generate the LHCII trimer starting structure, we aligned the CG structure of the full length monomer consisting of 232 amino acids with each subunit of the crystal structure trimer (22) and combined them to one structure (for details see Section S3 in the SI). In addition, we added three PG lipids which are present in the crystal structure (22) at the interface between the monomeric subunits and coordinate the Mg^{2+} ion of CLA611. The employed tail combination of the PG lipid is 16:1(3t)-16:0 PG. Note that the CG model for the monomer was employed for all trimer simulations and in particular no additional interactions between the LHCII monomers were added.

Trimeric LHCII was also embedded in a POPC bilayer consisting of 282 lipids per leaflet (box size 15.0×15.0×10.0 nm^3) to investigate the behavior in a model membrane. The protein-bilayer-system was solvated by ~10,100 CG water beads, neutralized, and 0.15 M NaCl was added. CG simulations of 3 μs were used to investigate the impact of the model membrane (Sections 3.1 and 3.2)

2.2 Setup and validation of the coarse-grained model of monomeric LHCII

To simulate the LHCII complex with the CG force field Martini, first CG models for all cofactors are required. While a chlorophyll A model already exists (23), all other cofactors – namely chlorophyll B, lutein, neoxanthin, and violaxanthin – had to be parametrized beforehand. In doing so, we used atomistic simulations with the force field GROMOS 53a6 (13, 24) in decane as reference. Details are given in Section S1 of the SI.

With the cofactors at hand, a CG model of the protein with the embedded cofactors can be set up. Figure 1 provides structural details of LHCII including all cofactors. The amino acid protonation states were chosen according to pH 7. While histidine 120 was protonated because it forms a salt bridge with aspartate 111 in the lumenal loop, histidine 68 and 212 were not protonated (22). The latter two coordinate the Mg^{2+} ions of chlorophyll A 603 and A 614. The CG structure of the protein was generated using the program *martinize.py*. In the case of the cofactors, isolated CG structures were generated which were subsequently combined with the CG protein structure. For some chlorophyll tails, not all atoms have been resolved in the crystal structure. These were added in similar configurations as the resolved ones at the coarse-grained resolution. The minimization of the CG structures allowed these tails to adopt a relaxed structure (for details see section S3 in the SI).

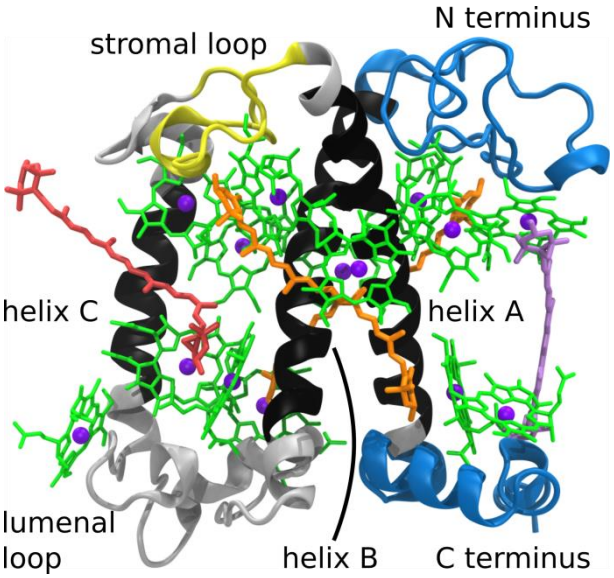

Figure 1. Structure of LHCII with all cofactors embedded in the protein. The aliphatic tails of the chlorophylls (green) are omitted for clarity, the central Mg^{2+} ions are shown as violet spheres. Neoxanthin is colored red, lutein orange, and violaxanthin light violet. The black regions of the three transmembrane helices A, B, and C are employed for root mean square deviation minimization before RMSF calculation.

We tested two different approaches to maintain the structural scaffold of the secondary and tertiary protein structure: first the commonly used elastic network dynamics (ElNeDyn) approach (25) and second the recently presented combination of Martini with a Gō-like model (26). In the Gō-like model, the interactions are described using Lennard-Jones potentials. Thus, it in principle allows for protein folding and unfolding. The structure-based contacts underlying the Gō-like model are derived from an overlap and contact of structural units-criterion (for details see ref. (26)). Contacts up to a distance of 1.1 nm were added which corresponds to the non-bonded cutoff distance. In total, 297 Lennard-Jones interactions are defined in the Gō-like model. To optimize the scaffold-maintaining models, atomistic simulations of the LHCII monomer with the force field GROMOS 54a7 in POPC membrane served again as a reference

(13). As the atomistic simulations contained only the residues 14-231 which are present in the crystal structure, the optimization of the elastic network was also performed with these. To obtain the final model, we modelled the missing residues with I-TASSER (27, 28) before coarse-graining them. Nevertheless, no additional interactions were introduced for the modelled residues in the ElNeDyn and Gō-like model, respectively, as the fact that the modelled residues were not resolved in crystal structure emphasizes their enhanced flexibility.

For both scaffold-maintaining models – ElNeDyn and Gō-like – we tested several parameter settings. The best parameters were chosen based on the comparison of the root mean square fluctuations (RMSF) of the Cα carbon atoms with the atomistic reference. The RMSF of the atomistic simulations were calculated for time intervals of 100 ns and averaged for three trajectories with a total simulation time of 3000 ns. Due to the speed up factor of ~3-10 in the CG simulations (6, 29), their RMSF were evaluated for time intervals of 30 ns. They were averaged for one trajectory of 900 ns.

Chlorophyll embeds a Mg^{2+} ion in the center of its aromatic moiety which is coordinated with four nitrogen atoms of which two are deprotonated. Thus, overall chlorophyll is neutral. In addition, coordinate covalent bonds of the Mg^{2+} with amino acid side chains, the protein backbone, other chlorophylls, or lipids are present in the crystal structure. Except for the lipid, these are modelled by additional Lennard-Jones potentials between the Mg^{2+} beads and the respective binding beads similar to the backbone bonds in the Gō-like model. The additional Lennard-Jones potentials are applied in both scaffold-maintaining models with a dissociation energy of $\varepsilon = 9.414$ kJ/mol. The bond to the lipid was omitted to allow the exchange with other lipids in the bilayer without pulling the chlorophyll out of the protein. In some cases, the bonds are mediated by water molecules (CLA604, CLB606, and CLB609). These water molecules

form a coordinate covalent bond to Mg^{2+} and hydrogen bonds to specific residues. Because this bonding situation is quite stable we bind these chlorophylls to the corresponding protein residue using a Lennard-Jones potential with $\varepsilon = 9.414$ kJ/mol. During most of the simulation time in the atomistic simulations used as a reference, water molecules are present to mediate the connection between the chlorophylls and the amino acids. Due to the CG level, the mediating water molecules are not represented explicitly as individual particles but their mediating function is taken into account. The good agreement of the chlorophylls' aromatic core RMSF between atomistic and CG LHCII monomers (cf. Figure S9 in the SI) supports the chosen value of the dissociation energy ε to represent the coordinate covalent bond. Note that the carotenoids do not form coordinate covalent bonds with the protein and thus no additional potential is added.

Figure 2a shows the RMSF of the atomistic simulations (blue line, UA) together with the two different CG models. The grey background indicates the regions of the three transmembrane α-helices (black regions of helix A, B, and C in Figure 1) which were used to minimize the root mean square deviations of the protein before calculating the RMSF. The magenta crosses mark the residues which form a coordinate covalent bond with the chlorophylls. For the ElNeDyn model, a force constant of $k = 500$ kJ/mol and a cutoff radius of $R_c = 0.8$ nm were used. For the Gō-like model, a dissociation energy of $\varepsilon = 9.414$ kJ/mol showed the best agreement with the atomistic RMSF. Overall, the RMSF of both CG models exhibit a good agreement with the atomistic reference.

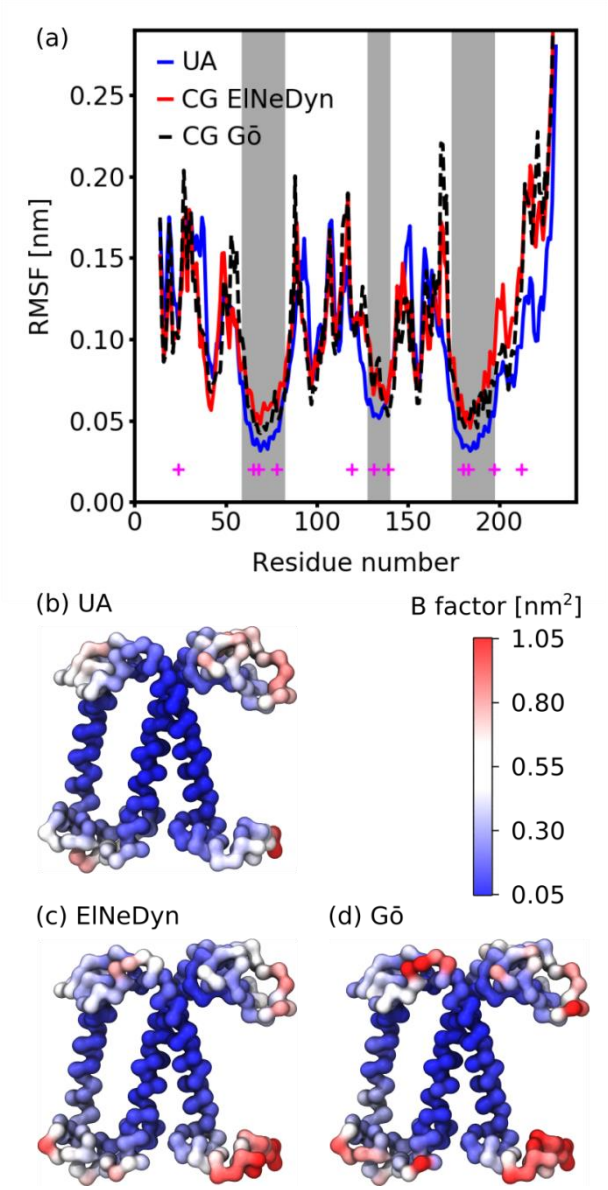

Figure 2. Comparison of the backbone RMSF of the different protein models. (a) LHCII monomer with the residues 14-231 resolved in the crystal structure simulated in a POPC bilayer. The RMSF of the atomistic Cα carbon atoms is shown in blue, the RMSF of the CG backbone beads in red (ElNeDyn) and dashed black (Gō), respectively. (b–d) Protein backbone of LHCII colored according to the B factor derived from the Cα/backbone RMSF of the various MD simulations.

Figures 2b–d show the backbone of the LHCII monomer colored according to the B factor calculated from the backbone RMSF in the MD simulations for the atomistic (b), CG ElNeDyn (c), and the CG Gō-like model (d), respectively. In all three cases, the transmembrane helices A, B, and C exhibit the lowest B factor. The loops and termini located at the membrane/water interface show a slightly higher flexibility in the CG models compared to the atomistic simulations. This is particularly the case for the C-terminal loop on the lower right side of the protein, as can also be seen in Figure 2a for residue numbers >210.

The results presented in the following Sections were all obtained using the Gō-like model because we decided to focus on this model for the description of the secondary and tertiary protein structure. The backbone and cofactor RMSF obtained with ElNeDyn are qualitatively identical emphasizing their robustness with respect to the protein model. They are presented in Section S5 of the SI.

3. RESULTS

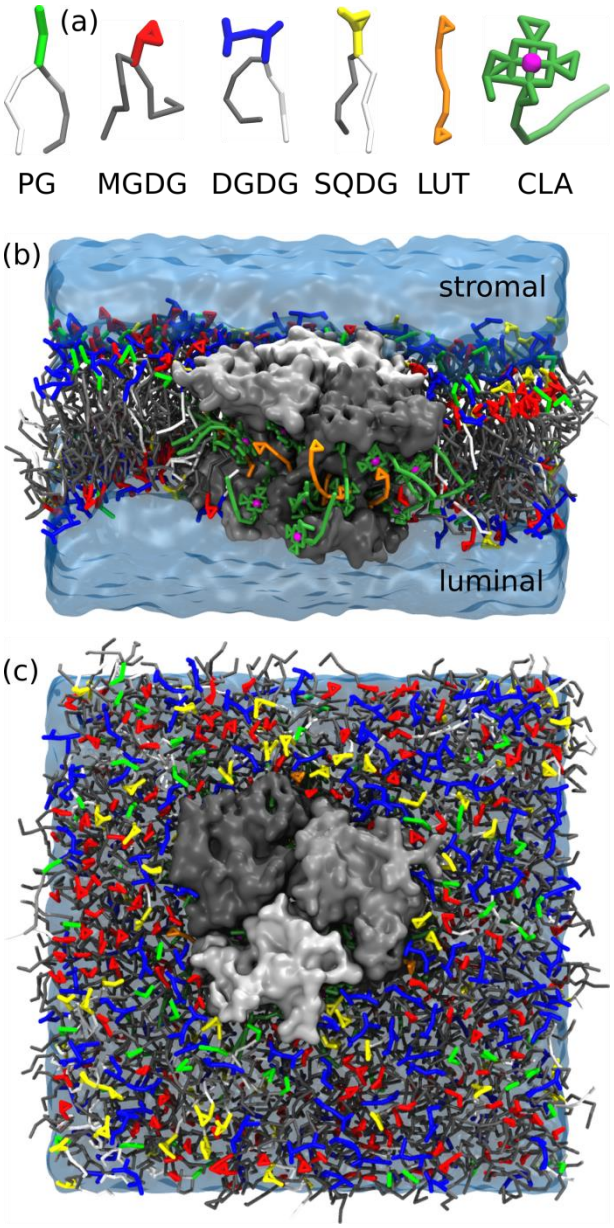

Figure 3. Details of the CG simulation setup. (a) CG representations of the four lipid classes present in the plant thylakoid membrane model, lutein (LUT) as carotenoid representative, and chlorophyll A (CLA) as chlorophyll representative. (b) Side and (c) stromal view of trimeric LHCII embedded in a thylakoid membrane patch. The LHCII monomer units are shown as grey surfaces, water as transparent blue surface. Lipids and cofactors are colored as in (a); all carotenoids are depicted in orange, all chlorophylls in dark green.

3.1 Dynamics of LHCII trimers

We performed simulations of trimeric LHCII in plant thylakoid membrane which is characterized by a high degree of tail unsaturation (see Figure 3). Approximately 30% of the lipids are negatively charged (PG, SQDG) and 85% of the total lipids contain a sugar headgroup (MGDG, DGDG, SQDG) (19–21). Figure 4 summarizes the properties of the LHCII trimer. To evaluate the stability of the protein complex, we analyze the temporal evolution of the monomer distances based on the two crossing transmembrane helices A and B for 30 μs of simulation. Figure 4a depicts the distances of their backbone centers of mass which change only slightly during the first 10 μs of the simulation. The three monomers show a steady average distance of 3.53±0.01 nm (3.69±0.01 nm / 3.63±0.01 nm / 3.26±0.01 nm) in the last 20 μs of the simulation which is in good agreement with the average distance in the crystal structure of 3.50 nm. Although the threefold symmetry gets slightly distorted, the overall crystal arrangement of the protein complex is stable in the thylakoid membrane on the μs time scale.

The residues 1-13 of the N-terminus which are not resolved in the crystal structure are preferentially located on top of the membrane-protein interface where they can interact with the lipid headgroups. This observation differs from recent position modelling of the N-terminus based on electron paramagnetic resonance (EPR) data (30). The discrepancy might be caused by different environments (thylakoid membrane versus *n*-dodecyl *β*-D-maltoside/sucrose/ glycerol), increased hydrophobicity due to the mutation of serine residues and the spin label, uncertainty of the rotamer configurations in EPR based structure modelling (31), or a potential overestimation of the conformational flexibility of the N-terminus in the present work due to missing supportive bonds in the residues which are not resolved in the crystal structure.

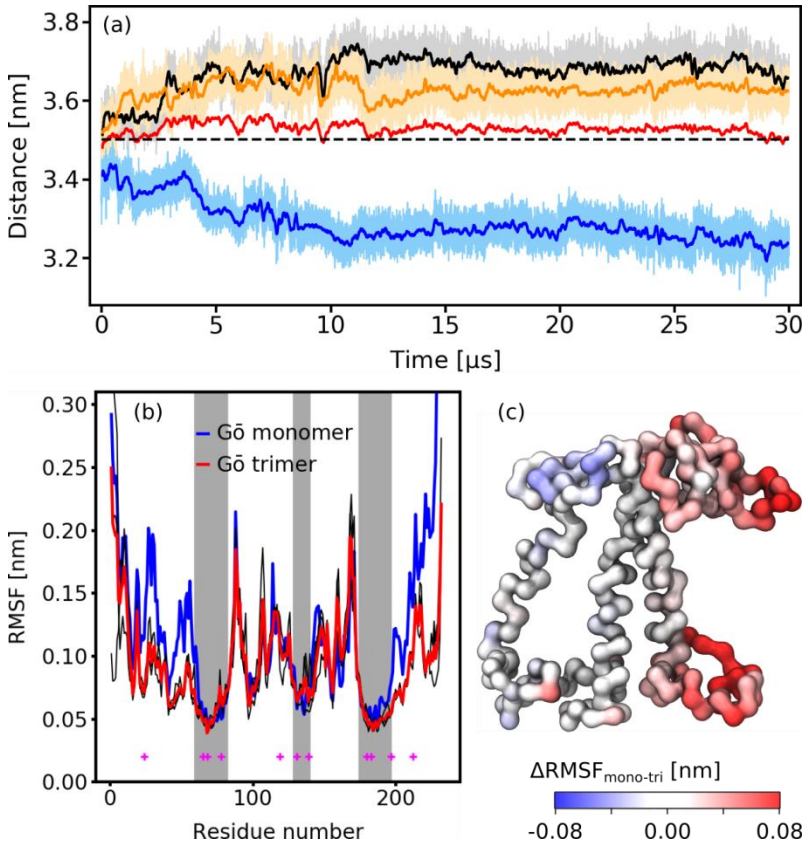

Figure 4. Properties of trimeric LHCII. (a) Monomer distances evaluated based on the backbone center of mass of the two crossing transmembrane helices. Individual monomer distances are colored grey, orange, and blue; the thicker line depicts the 100 ns running average. The thick red line shows the average monomer distance and the dashed black line indicates the distance in the crystal structure. (b) Effect of the trimer formation for the full length of LHCII (232 residues) in thylakoid membrane. The thin black line depicts the RMSF of each monomer within the trimer individually; the red line is their average; the blue line for the monomer simulation. (c) Protein backbone of LHCII colored according to the changes in RMSF from trimer to monomer in thylakoid membrane.

To clarify the impact of trimer formation, we compare the dynamics of the protein backbone of the full length trimer with the full length monomer. Figure 4b displays the RMSF of the protein

backbone of the monomer (blue line), the individual monomers in the trimeric complex (thin black lines), and the average of the monomers in the trimer (red line). The RMSF is calculated for 30 windows of 30 ns each evaluating a total time of 900 ns. Primarily, the monomer formation results in a rigidity loss of the regions with higher RMSF, the loops at the membrane-water interface. The transmembrane helices (grey shaded areas) however keep a similar stability as in the trimer. The reduced flexibility of the loops is even more visible in Figure 4c depicting the change in RMSF from trimer to monomer. Apparently, the N- as well as the C-terminus (residues 1–61 and 199–232; blue regions in Figure 1), which are located at the protein-protein interfaces in the LHCII trimer, gain flexibility. In contrast, the coil region between residues 152–170 in the stromal loop (yellow region in Figure 1) at the trimer periphery becomes slightly more stabilized in the monomer. This stabilization might result from a more rigid conformation which might be energetically more unfavorable in the trimer.

So far, we analyzed the behavior of LHCII in its natural environment, the thylakoid membrane. Changing the membrane to a simple POPC bilayer which exhibits a higher degree of saturation does lead to a slight decrease of the monomer distances in the trimer with an average of 3.38±0.01 nm. However, no significant changes of the protein backbone RMSF emerge. Figure S3 shows the temporal evolution of the monomer distances and Figure S4 compares the RMSF in both environments which are almost identical.

3.2 Dynamics of the cofactors

Besides the stability of the protein complex, the stability of the cofactors is also of interest as they are the key players for the function of the antenna complex: capturing photons and guiding them towards the reaction center. The cofactor dynamics are analyzed for the last 2 μs of a 3 μs

simulation. Figure 5 depicts the position of the chlorophylls in the LHCII trimer viewed from the stromal (a) and lumenal side (b) as well as the RMSF of their aromatic cores and tail ends averaged over 40 windows of 50 ns each for a simulation in the thylakoid membrane. In Figure 5a and b the spheres indicating the location of the chlorophylls in the LHCII trimer are colored according to their tail end RMSF. For both sides – stromal (a) as well as lumenal (b) – the tail ends of the chlorophylls in the protein center show a lower RMSF than the ones of the chlorophylls in the protein periphery. Moreover, the ones on the lumenal side have a higher RMSF than the ones on the stromal side. Figure 5c depicts that the aromatic cores (red circles) show a RMSF of 0.04–0.18 nm. According to the RMSF of the tail ends (blue squares), two sub-groups emerge: the first one exhibits similar tail RMSF as their cores (0.08–0.10 nm) and the members of this sub-group (CLA602, CLA603, CLB601, CLB607, and CLB609) are located towards the center of the trimer. The second sub-group possesses a tail end RMSF which is about fivefold higher (0.28–0.91 nm) than the RMSF of their aromatic cores. These chlorophylls are located at the protein-membrane interface.

The co-crystallized PG lipids located at the interface between the monomeric units of the LHCII trimer show no exchange event during the entire simulation time of the trimer (82 μs). The RMSF of the headgroup, which coordinates the Mg^{2+} of CLA611, is with 0.089±0.002 nm similar to the aromatic cores of the chlorophylls. The tail ends of the fatty acids show higher RMSF of 0.167±0.008 nm.

Let us now take a look at the chlorophyll flexibility in the LHCII monomer. In case of the aromatic cores only some exhibit a slight increase in flexibility compared to the trimer while most of them keep a similar RMSF in the monomer. This is to be expected as they form a coordinate covalent bond with the protein which is included in our CG model by an additional

Lennard-Jones potential. In contrast thereto, the tail ends show a clear loss of rigidity. The grey squares in Figure 5c show that particularly the chlorophylls from sub-group one, but also e.g. CLA612 and CLA613, become much more flexible. The correlation of the chlorophyll RMSF in trimeric and monomeric LHCII (Figure 5d) shows that about two thirds of the tail ends (blue squares) pronouncedly loose rigidity in the monomer; the others keep their flexibility. Overall, the chlorophyll RMSF emphasizes the reduced rigidity due to the removal of the trimeric interfaces.

Figure S9 provides a comparison between the atomistic and CG chlorophyll RMSF of monomeric LHCII in a POPC bilayer. The aromatic cores (dots) show a good agreement supporting the chosen strength of the added interaction to model the coordinate covalent bond of the Mg^{2+} ions. The RMSF of the tail ends are between about 0.2 – 1.0 nm for both systems.

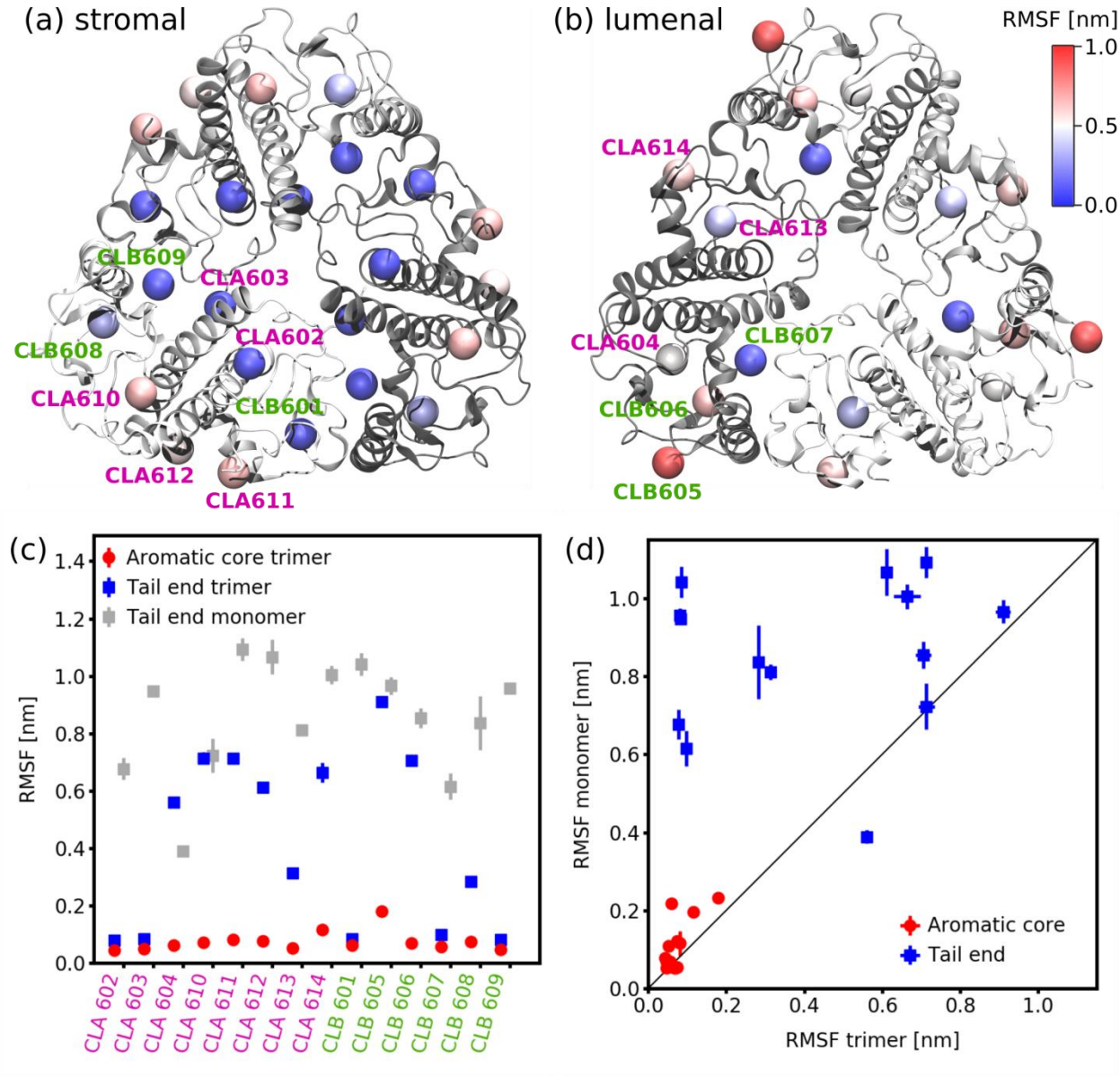

Figure 5. RMSF of the chlorophylls in LHCII. (a/b) Stromal and lumenal view of the Mg^{2+} positions (spheres) in the crystal structure of LHCII trimers. The spheres are colored according to the RMSF of the tail ends of the corresponding chlorophyll in a CG trimer simulation in thylakoid membrane. (c) RMSF of aromatic cores (red circles) and tail ends (blue squares) in LHCII trimers. The grey squares depict the RMSF of the chlorophyll tail ends in monomeric LHCII. (d) Correlation of the chlorophyll RMSF in trimeric and monomeric LHCII.

Figure 6 shows the impact of the membrane environment on the chlorophyll RMSF in trimeric LHCII. Overall, the impact of replacing the thylakoid membrane with a pure POPC bilayer is less pronounced than the impact of the monomer formation. The RMSF of the cores are almost unchanged (red circles in Figure 6b) whereas four of the tail ends become slightly less flexible in POPC bilayers by a ΔRMSF of approximately 0.1 nm. Again, these chlorophylls are located in the periphery of the trimer which enables them to easily interact with the lipid surrounding. The RMSF of the co-crystallized lipid headgroup also remains almost unchanged (0.070±0.001 nm) whereas its tail ends slightly gain flexibility (0.214±0.013 nm).

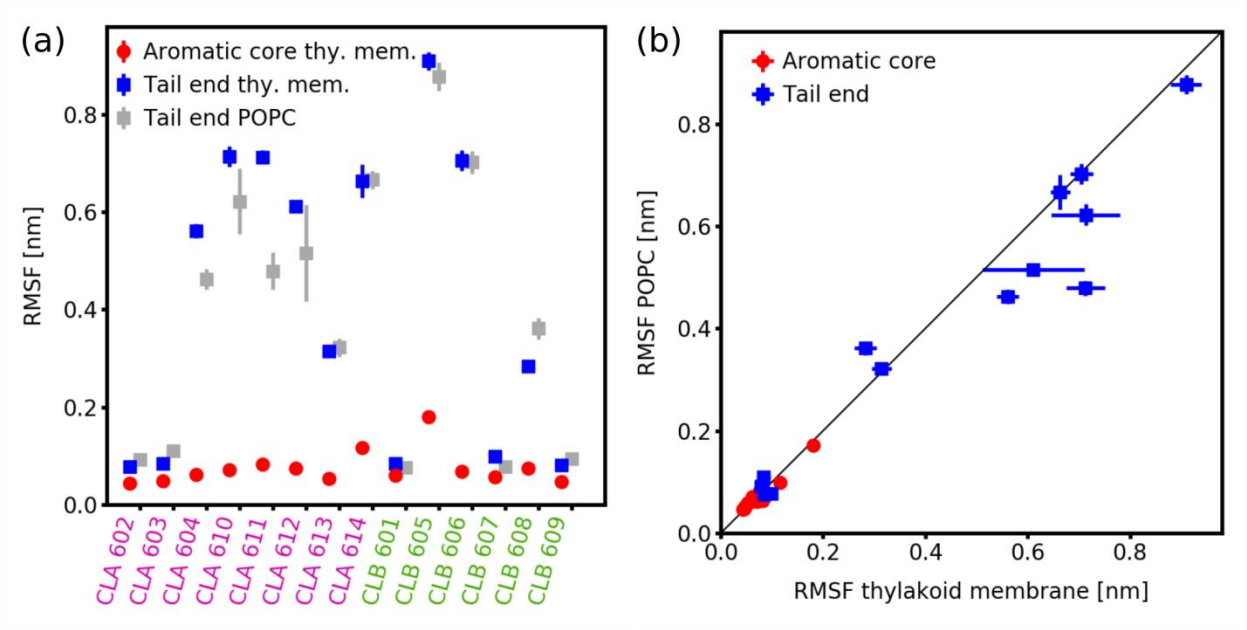

Figure 6. Impact of the membrane environment on the chlorophyll RMSF in LHCII. (a) RMSF of aromatic cores (red circles) and tail ends (blue squares) for LHCII trimers in thylakoid membrane. The grey squares depict the RMSF of the chlorophyll tail ends in trimeric LHCII embedded in a POPC bilayer. (b) Correlation of the chlorophyll RMSF of trimeric LHCII in thylakoid membrane versus POPC.

The efficiency of the excitation energy transfer between two chlorophylls is strongly dependent on the distance of the respective chlorophylls. Our simulations provide insights about how these distances changes upon monomer formation. Figure S10 (top) in Section S9 of the SI provides the distance matrix of the chlorophylls within one monomeric unit in the LHCII trimer (lower triangle) as well as the distance matrix within an isolated LHCII monomer in thylakoid membrane. The distances are measured for the Mg^{2+} ions. Most of the changes are within ±1 nm with a maximum of about 1.5 nm. On average, the chlorophyll distances increase by ~0.11 nm in the monomer compared to the trimer. The higher chlorophyll distances lead to decreased excitonic couplings. This highlights the importance of the trimeric state for the required compactness of the chlorophylls to perform the highly efficient energy transfer in LHCII.

Figure S12 (top) provides a comparison of the CG LHCII monomer (lower triangle) and the atomistic monomer (upper triangle) in a POPC bilayer. The distances increase in the CG simulation compared to the atomistic simulation by an average of ~0.12 nm. This increase might be mainly due to the more coarse representation of the chlorophylls as well as the amino acid sidechains in the CG simulation.

3.3 Thylakoid lipid fingerprints of LHCII

As the thylakoid membrane consists of seven different lipids, we analyzed the lipid surrounding of the LHCII trimer – its lipid fingerprint (9). The lipid densities were evaluated from two simulations of 30 μs each. The first 5 μs were discarded as equilibration process. Figure 7 depicts the lipid densities around LHCII resolved according to the four different headgroups present in thylakoid membrane: PG, MGDG, DGDG, and SQDG (from left to right). First of all, the high density of PG lipids within the protein (white area) is eye-catching (top left).

This arises from the three co-crystallized PG lipids embedded at the interface between the monomeric subunits. In the periphery of the LHCII trimer, the negatively charged PGs are strongly depleted on the stromal as well as the lumenal side. This is supported by the overall charge of −18 of a trimeric LHCII including the three co-crystallized PG lipids.

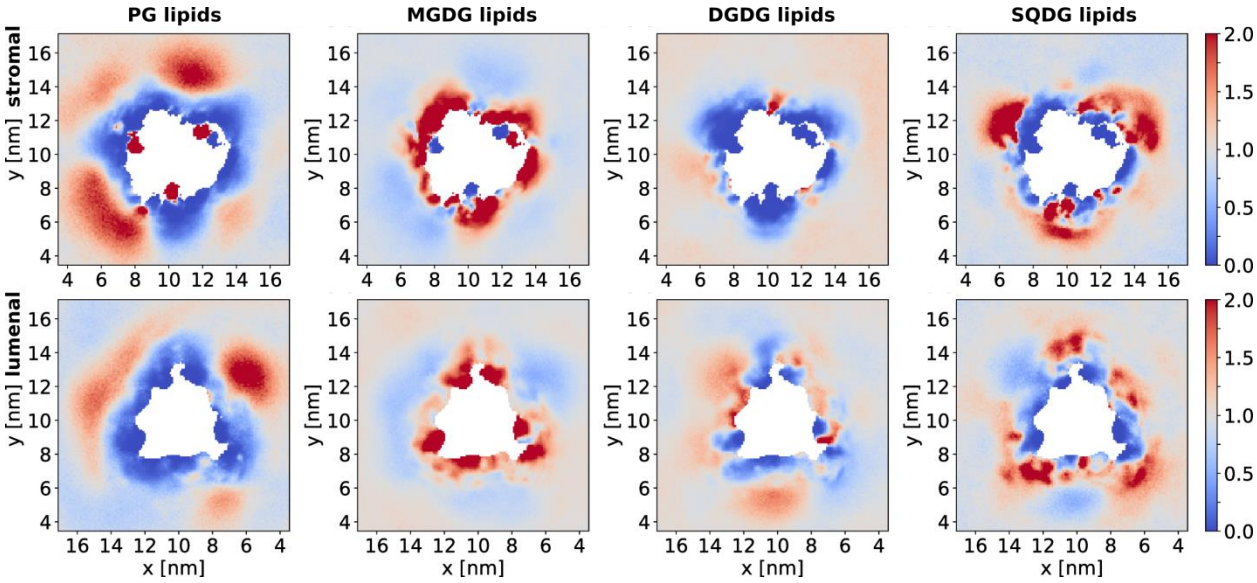

Figure 7. Lipid density maps around LHCII trimers analyzed depending on the four different headgroups present in plant thylakoid membrane. The top row depicts the stromal view; the bottom row the lumenal view. To obtain the lumenal view, the membrane was rotated around the y axis. Thus, both sides are depicted using a viewpoint in the water phase. The white area in the center represents the area covered by LHCII with a lipid density < 2% of the bulk membrane. The colorbar depicts the relative lipid density in relation to its fraction in the CG thylakoid membrane model.

In contrast to PGs, the MGDG lipids show a strong preference to the vicinity of LHCII (second column in Figure 7). While on the stromal side binding is observed throughout the protein surface, on the lumenal side six hot spots – two per LHCII monomer – are present. Moreover, a pronounced MGDG density joins the one of the PG densities at the monomer-monomer

interfaces on the stromal side. The overall strong affinity of MGDG to LHCII is in line with an experimental study suggesting that the galactolipids present in the thylakoid membrane – MGDG and DGDG – promote the formation of larger LHCII aggregates (32). Our MGDG density maps propose that MGDG stabilizes aggregates by mediating the interaction between multiple LHCII trimers. Recent atomic force microscopy studies suggest in addition that MGDG stabilizes LHCII against unfolding (33). In a mixed DGDG/MGDG membrane the unfolding force of LHCII trimers increased significantly compared to pure DGDG or POPG membranes which the authors attribute to the hourglass shape of LHCII and the good steric interplay with MGDG (cf. also Figure 8). Our simulations confirm that LHCII prefers MGDG over the other major thylakoid lipids in its surrounding. The preference for MGDG lipids was also observed for the cyanobacterial PSII core complex (12).

Whereas LHCII shows a strong preference to MGDG lipids this is not the case for DGDG lipids (third column in Figure 7). On the stromal side they are almost everywhere depleted on the protein surface. The lumenal side of LHCII shows some areas where DGDG lipids are enriched but it is much less pronounced than in the case of MGDGs. Although both – MGDG as well as DGDG – are glycolipids, the bulkiness of their sugar headgroups is quite different. While DGDG is a bilayer forming lipid, pure MGDG does not form bilayers. We attribute the preference of LHCII for MGDG mainly to the cone shape of MGDG which fits well with the shape of the protein (33). Figure 8 shows the volume density of the LHCII trimer and its cofactors (blue surface) averaged over 60 μs. The higher spatial demand of the protein on the top and bottom which corresponds to the head group region of the membrane is clearly visible. A second contribution to the preference of MGDG over DGDG could be of entropic nature due to the higher flexibility of the free DGDG head group compared to the free MGDG head group (12).

Comparable to PGs, the SQDG lipids are depleted in the immediate surrounding of LHCII. Again, the reason might be that SQDG is negatively charged. However, the distance range in which they are depleted is lower compared to PGs and there are clear hot spots present on both sides of the membrane. They have a distance of about 1–2 nm from protein core represented by the white regions.

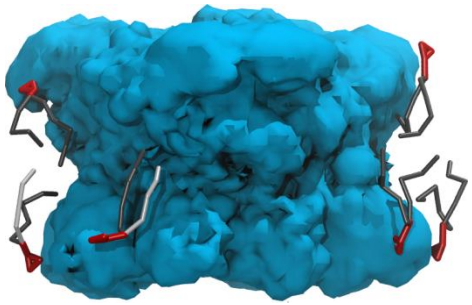

Figure 8. Averaged volume density of a LHCII trimer (blue) including its cofactors simulated in a thylakoid membrane. Six selected snapshots of MGDG lipids are shown.

The thylakoid lipid densities around LHCII trimers are in good agreement with experimental studies at mild detergent conditions which quantify the remaining lipids associated with isolated trimeric LHCII as being approximately 52% MGDG, 24% DGDG, 7% SQDG, and 13% PG (see Table 1) (34). The percentage PG corresponds to the three PG lipids found in the crystal structure at the trimeric interfaces. We evaluated the composition of the annular thylakoid lipid shell within a distance of 0.7 nm around the LHCII surface in our simulations. The observed composition is 57% MGDG, 18% DGDG, 15% SQDG, and 10% PG.

Table 1. Composition of the annular lipid shell around LHCII trimers. The relative enrichment compared to the membrane composition is given in brackets.

	PG	MGDG	DGDG	SQDG
simulation	10% (0.67)	57% (1.43)	18% (0.60)	15% (1.00)
experiment (34)	13% (0.82)	52% (1.27)	24% (0.97)	7% (0.70)

Besides the promotion of larger LHCII aggregates (32), the presence of MGDG has also been shown to affect the energy transfer from LHCII to PSII core complexes in proteoliposomes of thylakoid lipids (35). Separately isolated PSII core complexes and LHCII trimers have been reconstituted into proteoliposomes with and without MGDG. With MGDG the fluorescence of the PSII core complex as well as the photochemical activity increased significantly indicating a more efficient energy transfer due to a stronger interaction of the PSII core with the LHCII antennae.

Figure S8 in the SI depicts the lipid density maps of the individual lipids around LHCII trimers. A closer view reveals that particularly di-18:3(9,12,15) MGDG is enriched in the close vicinity of LHCII trimers. LHCII prefers the MGDG with two polyunsaturated tails (di-18:3(9,12,15) MGDG) over MGDG with only one polyunsaturated tail (18:3(9,12,15)-16:0 MGDG). This provides additional support to the conclusion drawn from Figure 8, namely that the protein shape is the main reason for the preference of MGDG because the two polyunsaturated lipid tails enhance the cone shape of MGDG. However, the lipid tails only play a role as secondary selection criterion (see also Section S7 in the SI).

6. CONCLUSIONS AND OUTLOOK

In summary, we presented a new coarse-grained model of light-harvesting complex II, the most important antennae complex for plant photosynthesis. Our CG simulations emphasize the

importance of trimer formation to stabilize protein loops at the membrane-water interface as well as chlorophylls in the center of the protein complex. Although the minor light-harvesting proteins CP24, CP26, and CP29, which are closely related to LHCII, occur as monomers, their embedding in the PSII supercomplex results in several close contacts to neighboring proteins (2, 3). Despite the different structures of the protein-protein interfaces, we expect these interactions to stabilize the minor light-harvesting proteins in a similar fashion as LHCII monomers are stabilized in the trimer.

The impact of the membrane environment on the protein and cofactor dynamics is less pronounced than the trimer formation. Trimeric LHCII shows a clear preference to MGDG lipids in its surrounding which is in good agreement with experimental observations at mild detergent conditions (34). Recently, van Eerden et al. have shown that MGDG and SQDG are enriched in the annular shell of a cyanobacterial photosystem II core complex (12). Together with our observed preference of LHCII for MGDG, this strongly supports observations that MGDG has a distinct impact on the absorption spectra, the energy transfer efficiency, and the oxygen evolution in reconstituted mixtures of PSII core complexes and LHCII antennae (35) by fostering the contacts between the involved protein complexes.

In addition, our simulations emphasize the chlorophylls' sensitivity towards changes in the protein-protein interactions which can be observed in their flexibility and distances. In turn, this gives the protein the possibility to impact its chromophores' behavior and thus its excitonic properties by changing their interactions with other proteins. This is particularly interesting as the regulation of the major antenna protein LHCII has a key importance for photoprotection of the photosynthetic machinery in plants. Two of the discussed mechanisms of photoprotection – LHCII aggregation and PsbS binding – involve changes in the protein-protein interactions of

LHCII (36). Based on the presented results we expect that the interaction of LHCII trimers with additional proteins leads to changes in the chlorophyll flexibility which potentially open additional dissipation channels for the excitation energy. Moreover, our protein model opens a way to study in detail LHCII oligomerization as well as interaction with other protein complexes involved in energy transmission to the reaction center or excitation energy quenching.

In the future, we also aim to employ our new well-tuned CG model to study the diversity of chromophore arrangements in LHCII on the hundreds of μs time scale. In particular the chlorophylls are responsible for capturing photons and transmitting the excitation energy towards the PSII reaction center, a process taking place on the picosecond time scale. In contrast, the dynamics of the protein matrix in which the optically active chromophores are embedded covers a much longer time scale. This implies that the protein fluctuations do not drastically change one single excitation event but they generate a manifold of different conditions in terms of chromophore distances and orientations which are all present at the same time in chloroplasts. Our CG model of LHCII can help unravelling the degree of diversity of the chromophore arrangement and its impact on natural photosynthesis.

ASSOCIATED CONTENT

Supporting Information. The following files are available free of charge.

Details of the CG parametrization of the LHCII cofactors, simulation times, the procedure to build the CG structure of the monomer and trimer, the monomer distances as well as backbone RMSF of trimeric LHCII in POPC bilayers, the backbone and chlorophyll RMSF with the ElNeDyn model, the backbone RMSF of trimeric compared to monomeric LHCII in POPC

bilayers, the individual lipid densities, the chlorophyll RMSF of the atomistic simulations, and the chlorophyll distances. (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: s.thallmair@rug.nl

*E-mail: s.j.marrink@rug.nl

Author Contributions

S.T. and S.J.M. designed the research; S.T. and P.A.V. performed the research. The manuscript was written by S.T. with contributions of all other authors.

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

We thank Nicoletta Liguori and Roberta Croce for providing the atomistic trajectories of monomeric LHCII from reference (13) and Paulo C. T. Souza for fruitful discussions. We acknowledge financial support by the European Commission via an ERC grant to S.J.M. (COMP-MICR-CROW-MEM, grant agreement 669723), a Marie Skłodowska-Curie Actions Individual Fellowship to S.T. (MicroMod-PSII, grant agreement 748895), and an Erasmus+ Mobility Fund to P.A.V.. This work was supported via computing time from the National Computing Facilities Foundation (NCF) of The Netherlands Organization for Scientific Research (NWO).

REFERENCES

1. Blankenship, R.E. 2014. Molecular Mechanisms of Photosynthesis. Wiley VCH.

2. Wei, X., X. Su, P. Cao, X. Liu, W. Chang, M. Li, X. Zhang, and Z. Liu. 2016. Structure of spinach photosystem II–LHCII supercomplex at 3.2 Å resolution. Nature. 534: 69–74.

3. Su, X., J. Ma, X. Wei, P. Cao, D. Zhu, W. Chang, Z. Liu, X. Zhang, and M. Li. 2017. Structure and assembly mechanism of plant C2S2M2-type PSII-LHCII supercomplex. Science. 357: 815–820.

4. Ruban, A. V., M.P. Johnson, and C.D.P. Duffy. 2012. The photoprotective molecular switch in the photosystem II antenna. Biochim. Biophys. Acta - Bioenerg. 1817: 167–181.

5. Jahns, P., D. Latowski, and K. Strzałka. 2009. Mechanism and regulation of the violaxanthin cycle: The role of antenna proteins and membrane lipids. Biochim. Biophys. Acta - Bioenerg. 1787: 3–14.

6. Marrink, S.J., H.J. Risselada, S. Yefimov, D.P. Tieleman, and A.H. de Vries. 2007. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. J. Phys. Chem. B. 111: 7812–7824.

7. de Jong, D.H., G. Singh, W.F.D. Bennett, C. Arnarez, T.A. Wassenaar, L. V. Schäfer, X. Periole, D.P. Tieleman, and S.J. Marrink. 2013. Improved Parameters for the Martini Coarse-Grained Protein Force Field. J. Chem. Theory Comput. 9: 687–697.

8. Hedger, G., and M.S.P. Sansom. 2016. Lipid interaction sites on channels, transporters and receptors: Recent insights from molecular dynamics simulations. Biochim. Biophys. Acta - Biomembr. 1858: 2390–2400.

9. Corradi, V., E. Mendez-Villuendas, H.I. Ingólfsson, R.-X. Gu, I. Siuda, M.N. Melo, A. Moussatova, L.J. DeGagné, B.I. Sejdiu, G. Singh, T.A. Wassenaar, K. Delgado Magnero, S.J. Marrink, and D.P. Tieleman. 2018. Lipid–Protein Interactions Are Unique Fingerprints for Membrane Proteins. ACS Cent. Sci. 4: 709–717.

10. Corradi, V., B.I. Sejdiu, H. Mesa-Galloso, H. Abdizadeh, S.Y. Noskov, S.J. Marrink, and D.P. Tieleman. 2018. Emerging Diversity In Lipid-Protein Interactions. Chem. Rev. : 10.1021/acs.chemrev.8b00451.

11. van Eerden, F.J., T. van den Berg, P.W.J.M. Frederix, D.H. de Jong, X. Periole, and S.J. Marrink. 2017. Molecular Dynamics of Photosystem II Embedded in the Thylakoid Membrane. J. Phys. Chem. B. 121: 3237–3249.

12. Van Eerden, F.J., M.N. Melo, P.W.J.M. Frederix, and S.J. Marrink. 2017. Prediction of Thylakoid Lipid Binding Sites on Photosystem II. Biophys. J. 113: 2669–2681.

13. Liguori, N., X. Periole, S.J. Marrink, and R. Croce. 2015. From light-harvesting to photoprotection: structural basis of the dynamic switch of the major antenna complex of plants (LHCII). Sci. Rep. 5: 15661.

14. Abraham, M.J., T. Murtola, R. Schulz, S. Pálla, J.C. Smith, B. Hess, and E. Lindahl. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 1–2: 19–25.

15. de Jong, D.H., S. Baoukina, H.I. Ingólfsson, and S.J. Marrink. 2016. Martini straight: Boosting performance using a shorter cutoff and GPUs. Comput. Phys. Commun. 199: 1–7.

16. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. J. Chem. Phys. 126: 014101.

17. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. 52: 7182–7190.

18. Wassenaar, T.A., H.I. Ingólfsson, R.A. Böckmann, D.P. Tieleman, and S.J. Marrink. 2015. Computational Lipidomics with *insane* : A Versatile Tool for Generating Custom Membranes for Molecular Simulations. J. Chem. Theory Comput. 11: 2144–2155.

19. van Eerden, F.J., D.H. de Jong, A.H. de Vries, T.A. Wassenaar, and S.J. Marrink. 2015. Characterization of thylakoid lipid membranes from cyanobacteria and higher plants by molecular dynamics simulations. Biochim. Biophys. Acta. 1848: 1319–30.

20. Siegenthaler, P.A., and N. Murata. 1998. Lipids in photosynthesis : structure, function, and genetics. Springer.

21. Sakurai, I., J.-R. Shen, J. Leng, S. Ohashi, M. Kobayashi, and H. Wada. 2006. Lipids in oxygen-evolving photosystem II complexes of cyanobacteria and higher plants. J. Biochem. 140: 201–9.

22. Liu, Z., H. Yan, K. Wang, T. Kuang, J. Zhang, L. Gui, X. An, and W. Chang. 2004. Crystal structure of spinach major light-harvesting complex at 2.72 A resolution. Nature. 428: 287–92.

23. de Jong, D.H., N. Liguori, T. van den Berg, C. Arnarez, X. Periole, and S.J. Marrink. 2015. Atomistic and Coarse Grain Topologies for the Cofactors Associated with the Photosystem II Core Complex. J. Phys. Chem. B. 119: 7791–803.

24. Oostenbrink, C., T.A. Soares, N.F.A. van der Vegt, and W.F. van Gunsteren. 2005. Validation of the 53A6 GROMOS force field. Eur. Biophys. J. 34: 273–284.

25. Periole, X., M. Cavalli, S.-J. Marrink, and M.A. Ceruso. 2009. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. J. Chem. Theory Comput. 5: 2531–2543.

26. Poma, A.B., M. Cieplak, and P.E. Theodorakis. 2017. Combining the MARTINI and Structure-Based Coarse-Grained Approaches for the Molecular Dynamics Studies of Conformational Transitions in Proteins. J. Chem. Theory Comput. 13: 1366–1374.

27. Roy, A., A. Kucukural, and Y. Zhang. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5: 725–738.

28. Yang, J., R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang. 2015. The I-TASSER Suite: protein structure and function prediction. Nat. Methods. 12: 7–8.

29. Marrink, S.J., A.H. de Vries, and A.E. Mark. 2004. Coarse Grained Model for Semiquantitative Lipid Simulations. J. Phys. Chem. B. 108: 750–760.

30. Fehr, N., C. Dietz, Y. Polyhach, T. von Hagens, G. Jeschke, and H. Paulsen. 2015. Modeling of the N-terminal Section and the Lumenal Loop of Trimeric Light Harvesting Complex II (LHCII) by Using EPR. J. Biol. Chem. 290: 26007–20.

31. Jeschke, G. 2012. DEER Distance Measurements on Proteins. Annu. Rev. Phys. Chem. 63: 419–446.

32. Schaller, S., D. Latowski, M. Jemioła-Rzemińska, A. Dawood, C. Wilhelm, K. Strzałka, and R. Goss. 2011. Regulation of LHCII aggregation by different thylakoid membrane

lipids. Biochim. Biophys. Acta - Bioenerg. 1807: 326–335.

33. Seiwert, D., H. Witt, A. Janshoff, and H. Paulsen. 2017. The non-bilayer lipid MGDG stabilizes the major light-harvesting complex (LHCII) against unfolding. Sci. Rep. 7: 5158.

34. Schaller, S., D. Latowski, M. Jemioła-Rzemińska, C. Wilhelm, K. Strzałka, and R. Goss. 2010. The main thylakoid membrane lipid monogalactosyldiacylglycerol (MGDG) promotes the de-epoxidation of violaxanthin associated with the light-harvesting complex of photosystem II (LHCII). Biochim. Biophys. Acta - Bioenerg. 1797: 414–424.

35. Zhou, F., S. Liu, Z. Hu, T. Kuang, H. Paulsen, and C. Yang. 2009. Effect of monogalactosyldiacylglycerol on the interaction between photosystem II core complex and its antenna complexes in liposomes of thylakoid lipids. Photosynth. Res. 99: 185–193.

36. Ruban, A. V. 2016. Nonphotochemical Chlorophyll Fluorescence Quenching: Mechanism and Effectiveness in Protecting Plants from Photodamage. Plant Physiol. 170: 1903–16.